A framework for robustness to ambiguity of higher-order beliefs

Ronald Stauber

Accepted: 8 October 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract Standard type spaces induce belief structures defined by precise beliefs. This paper proposes and analyzes simple procedures for constructing perturbations of such belief structures in which beliefs have a degree of ambiguity. Specifically, we construct ambiguous type spaces whose induced (ambiguous) belief hierarchies approximate the standard, precise, belief hierarchies corresponding to the initial type space. Based on a metric that captures the resulting approximation, two alternative procedures to construct such perturbations are introduced, and are shown to yield a simple and intuitive characterization of convergence to the initial unperturbed environment. As a special case, one of these procedures is shown to characterize the set of all finite perturbations. The introduced perturbations and their convergence properties provide conceptual foundations for the analysis of robustness to ambiguity of various solutions concepts, and for various decision rules under ambiguity.

Keywords Bayesian games · Type spaces · Ambiguity · Robustness

JEL Classification C72 · D83

1 Introduction

Starting with the seminal papers of [Gilboa and Schmeidler \(1989\)](#) and [Bewley \(2002\)](#), an extensive literature has developed in economics that considers agents whose beliefs are ambiguous, in the sense that they can be represented by sets of probability measures, as opposed to precise beliefs that can be represented by unique measures. One

R. Stauber (✉)
Research School of Economics, Australian National University,
Canberra, ACT 0200, Australia
e-mail: ronald.stauber@anu.edu.au

important task in introducing models with ambiguous beliefs into the economics literature, is to establish to what extent results that were derived for models with precise beliefs still apply when agents' beliefs may be ambiguous. This paper provides an initial step for such a robustness analysis in the context of Bayesian games, by characterizing simple procedures that can be used to introduce "small amounts" of ambiguity to the interacting belief structures associated with such games.

Defining ambiguous perturbations of precise Bayesian beliefs is relatively straightforward in choice-theoretic settings where the uncertainty is one-sided. For example, if we consider an agent who is uncertain about the realization of a parameter θ which takes on values from a set Θ , a model with precise beliefs would represent the agent's beliefs using a probability measure p from the set $\Delta\Theta$ of probability measures with support in Θ . If the set $\Delta\Theta$ is endowed with a metric, a perturbation of p can be defined by some ε -neighborhood of p with respect to this metric.¹ An alternative approach to define ambiguous perturbations of beliefs that has recently gained popularity in the economics literature, is based on the notion of ε -contamination studied in Bayesian statistics.² Given any measure p defined over a set Θ , a set of measures $\Lambda \subseteq \Delta\Theta$, and any $\varepsilon \in (0, 1]$, an ε -contamination of p is defined by the set³

$$p_{\varepsilon|\Lambda} := \{q \in \Delta\Theta \mid q = (1 - \varepsilon)p + \varepsilon r, \text{ with } r \in \Lambda\}.$$

In Bayesian games, each player's relevant space of uncertain events includes not only the parameters describing the game, but also the possible beliefs of his opponents. Following [Harsanyi \(1967–1968\)](#), the resulting structures of interacting beliefs are modeled using the notion of a type space. Denoting the set of feasible parameters by Θ , a type space is defined by assigning to each player i a set of types S_i and a map $\eta_i : S_i \rightarrow \Delta(\Theta \times S_{-i})$ that associates to each type $s_i \in S_i$ a probability measure over the set of parameters Θ and the set of opponents' type profiles S_{-i} . The probability measure $\eta_i(s_i) \in \Delta(\Theta \times S_{-i})$ can then be interpreted as representing the (precise) beliefs of type s_i . Given the interacting nature of beliefs associated with Bayesian games, the assumption that agents' beliefs are precise is clearly a much stronger requirement for a game-theoretic environment than it is for an individual choice setting. Moreover, applying standard approaches based on metric neighborhoods or ε -contaminations to define ambiguous perturbations of the associated belief structures is problematic in the context of a game. To see this, note that the measure $\eta_i(s_i)$ associated to a type s_i in a type space model is an element of $\Delta(\Theta \times S_{-i})$, and therefore, any neighborhood of $\eta_i(s_i)$ and any ε -contamination of $\eta_i(s_i)$ is identified by a subset of $\Delta(\Theta \times S_{-i})$. If we view such subsets as representing the ambiguous beliefs of a player, such sets essentially define ambiguous beliefs over precise

¹ As an example, [Bergemann and Schlag \(2011\)](#) use the Prohorov metric to perturb beliefs in their study of robust monopoly pricing.

² See, for example, [Berger \(1985\)](#) or [Wasserman and Kadane \(1990\)](#).

³ Such ε -contaminated beliefs have been used in an auction context by [Bose et al. \(2006\)](#) and [Bose and Daripa \(2009\)](#), in a search model by [Nishimura and Ozaki \(2004\)](#), and are also mentioned as an example for a general robust mechanism design problem in [Lopomo et al. \(2009\)](#). Decision-theoretic foundations for the use of ε -contaminated beliefs are provided by [Nishimura and Ozaki \(2006\)](#) and [Kopylov \(2008\)](#).

beliefs, since they contain measures over Θ and S_{-i} , and each profile $s_{-i} \in S_{-i}$ is associated with a profile of precise beliefs for the opponents of player i . Thus, unless we re-interpret each s_i as a representation of ambiguous beliefs,⁴ using metric neighborhoods or ε -contaminations to define ambiguous perturbations seems to lead to inconsistent models where players are allowed to consider the possibility that their own beliefs may be ambiguous, but are still required to believe that their opponents' beliefs are precise.

It follows that a consistent model of an ambiguous perturbation should be based on a belief structure that simultaneously perturbs both a type's beliefs and the support of these beliefs. This would require players to be endowed with ambiguous beliefs, which also assign positive probability to the players' opponents having ambiguous beliefs. Such belief structures can be represented by ambiguous type spaces, where types are implicitly identified with sets of probability measures over the underlying parameter space and the set of (ambiguous) types of each player's opponents.⁵ Using the concept of an ambiguous type space, we can define an ambiguous perturbation of a precise non-ambiguous type space as an ambiguous type space whose belief structure is in some way "close" to the belief structure of the initial non-ambiguous type space. This paper explores how such a notion of closeness could be defined, and analyzes associated perturbations.

To define when an ambiguous type space can be viewed as approximating the belief structure associated with an initial non-ambiguous type space, we consider the standard embedding of such type spaces into the ambiguous universal type space, which was characterized by [Ahn \(2007\)](#), who extended analogous constructions of the non-ambiguous universal type space as introduced by [Mertens and Zamir \(1985\)](#) and [Brandenburger and Dekel \(1993\)](#). In the ambiguous universal type space, types are defined by all infinite hierarchies of ambiguous beliefs about parameters, ambiguous beliefs about opponents' ambiguous beliefs about parameters, and so on ad infinitum. Every such belief hierarchy that satisfies a consistency requirement defines a type in a corresponding type space, which is universal in the sense that every ambiguous type space can be embedded into this type space in a belief-preserving way. This embedding shows that any type space implicitly assigns to each type a corresponding hierarchy of beliefs and higher-order beliefs. Thus, given a topology on the belief hierarchies comprising the universal type space, we can measure how well an ambiguous type space approximates a given non-ambiguous type space by comparing the induced sets of belief hierarchies. We use the product topology defined by the weak topology over probability measures associated to the individual levels of the belief hierarchies as a notion of closeness of belief hierarchies, and define a corresponding metric on the set of type spaces that metrizes the distance between the sets of belief hierarchies induced by two type spaces. Intuitively, we consider a researcher who wants to analyze a strategic situation based on a non-ambiguous beliefs structure, even though the players of the game may only be able to form ambiguous beliefs. He will then want to insure that the

⁴ We will discuss later how this can be achieved in a mathematically precise way.

⁵ Ambiguous type spaces were introduced in the Bayesian games literature by [Kajii and Ui \(2005\)](#), and were further analyzed by [Ahn \(2007\)](#).

predictions he derives from his model are not too different from those that would arise if the possibly ambiguous belief hierarchies that the players may possess are close to those specified by his model. We will discuss the use of the product topology and the interpretation of our metric on type spaces more precisely when we introduce the corresponding formal definitions.

Based on the introduced metric on type spaces, we define a perturbation of a non-ambiguous type space as an ambiguous type space that firstly, contains for each non-ambiguous type at least one ambiguous type that approximates it using the product metric on induced belief hierarchies, and secondly, does not contain any ambiguous types whose distance (using the same metric) to the set of non-ambiguous types exceeds a given bound. While this definition seems to correctly identify when two belief structures can be considered as being close, it provides no useful procedure to derive associated perturbations. Moreover, using belief hierarchies to characterize the convergence of a sequence of perturbations to the initial unperturbed environment is impractical from an applied viewpoint. We show how these issues can be resolved by introducing two alternative methods to construct perturbations that allow a simple characterization of convergence to the initial belief structure.

We start with a procedure that defines an approximating ambiguous type space based on finite partitions of the initial non-ambiguous type space. The associated construction, which essentially identifies each partition element with an ambiguous type, was introduced in [Stauber \(2011\)](#). The resulting procedure to perturb the initial belief structures by considering finite partitions with small mesh is consistent with the standard partitional model of knowledge, as we can think of players only being able to identify the partition elements that contain their “true types.” However, if we view an abstract type space as an implicit representation of the players’ belief hierarchies, it is not clear whether the mesh of the partition that defines such an approximation is an appropriate measure for the distance between the belief hierarchies corresponding to the original types and their ambiguous counterparts. In the current paper, we construct the belief hierarchies induced by such partitional approximations, and prove that as the mesh of the partitions converges to zero, the resulting ambiguous belief hierarchies converge, in the product topology induced by the weak topology on the individual belief levels, to the belief hierarchies defined by the initial non-ambiguous types. This result shows that the mesh of the partition used to define a partitional type space is an appropriate measure for the degree of the associated approximation, and thus provides a simple and intuitive procedure that can be used to construct ambiguous perturbations that are consistent with the general definition we introduce, but do not require the analyst to consider the induced belief hierarchies.

While partitional type spaces provide an easy way to define ambiguous perturbations, they may not be appropriate in every situation, such as for example, when the initial type space is finite. However, we show that the conceptual approach taken in the definition of a partitional type space can be extended to define a more general class of perturbations, which includes all finite ambiguous perturbations, and also yields as special cases perturbations that generalize ε -neighborhoods and ε -contaminations to environments with interacting beliefs. Moreover, the degree of the approximation resulting from such perturbations, and the associated convergence properties, can be

characterized using a measure that is analogous to the mesh of the partition defining a partitional type space.

By introducing simple and tractable procedures to define and analyze ambiguous perturbations of interacting belief structures associated with Bayesian games, our paper establishes a foundation and a framework for analyzing the robustness of solution concepts to the introduction of ambiguous beliefs. [Stauber \(2011\)](#) applies such a framework to analyze the robustness of Bayesian Nash equilibria in an environment where players use decision rules under ambiguity as introduced by [Bewley \(2002\)](#). However, the methods proposed in the current paper are equally applicable for other solution concepts and alternative decision rules under ambiguity.

The paper is structured as follows: A model of ambiguous type spaces and their embedding into the universal ambiguous type space is presented in Sect. 2. Section 3 proposes a metric on the set of such type spaces, and introduces a formal definition of an ambiguous perturbation. Section 4 analyzes ambiguous perturbations defined by finite partitions of the initial type space, and Sect. 5 builds on this approach to introduce and characterize a general class of perturbations.

2 Type spaces and higher-order beliefs

We will restrict our analysis to compact metric sets. Given any compact metric space Y , and any Borel subset $E \subset Y$, we use $\text{cl}(E)$ to denote the closure of E , ΔY to denote the set of probability measures with support in Y , and $\mathcal{K}Y$ to denote the set of all non-empty compact subsets of Y . Note that if Y is a compact metric space, then ΔY is compact using the weak topology, and can be metrized using the Prohorov metric or any equivalent metric ([Dudley 2002](#), Chap. 11). Similarly, $\mathcal{K}Y$ is compact using the Hausdorff metric ([Kechris 1995](#), p. 26), which also implies that $\mathcal{K}\Delta Y$ is compact using the Hausdorff metric derived from the Prohorov metric. For any measurable function $f : Y \rightarrow Z$ and $\mu \in \Delta Y$, the image measure of μ induced by f is given by the map $\mathcal{L}_f : \Delta Y \rightarrow \Delta Z$ defined by $[\mathcal{L}_f(\mu)](E) := \mu(f^{-1}(E))$ for any Borel set $E \subset Z$. Furthermore, if $f : Y \rightarrow Z$ is continuous, we define its extension $f^{\mathcal{K}} : \mathcal{K}Y \rightarrow \mathcal{K}Z$ by $f^{\mathcal{K}}(E) := f(E)$. It is a standard result that continuity of f implies the continuity of $f^{\mathcal{K}}$ and $\mathcal{L}_f$, and thus of $\mathcal{L}_f^{\mathcal{K}}$.⁶

Given a finite set of players N and a compact set of parameters Θ , let $(S, \eta) \equiv (S_i, \eta_i)_{i \in N}$ denote a Θ -based non-ambiguous type space, where each S_i is a compact metric space, and each η_i is a continuous map $\eta_i : S_i \rightarrow \Delta(\Theta \times S_{-i})$ that defines a homeomorphism between S_i and its image in $\Delta(\Theta \times S_{-i})$, endowed with the weak topology.⁷ Let d^P denote the Prohorov metric on $\Delta(\Theta \times S_{-i})$, and define $d^*(s_i, s'_i) := d^P(\eta_i(s_i), \eta_i(s'_i))$, which implies that d^* is a metric on S_i that is equivalent to the metric defining S_i as a compact metric space (see [Dudley 2002](#), p. 40).

We will characterize a class of ambiguous perturbations of a non-ambiguous type space (S, η) using finite partitions of the type sets S_i . We denote such partitions by $(\pi_i)_{i \in N}$, and use $\pi_i(s_i)$ to refer to the partition element containing the type s_i . We

⁶ See also Ahn ([2007](#), pp. 287–288).

⁷ We use standard notation, so S_{-i} denotes the Cartesian product of all S_j with $j \neq i$.

assume that each partition element is measurable, in the sense that every element is a Borel set with respect to the metric topology on S_i , and define the mesh of a partition $\pi := (\pi_i)_{i \in N}$ as

$$\mathcal{M}(\pi) := \max_i \sup_{s_i} \sup_{s'_i, s''_i \in \pi_i(s_i)} d^*(s'_i, s''_i).$$

To define an ambiguous type space, the single-valued maps η_i are replaced by maps γ_i with values given by compact sets of probability measures. Denote ambiguous type spaces by $(T, \gamma) \equiv (T_i, \gamma_i)_{i \in N}$, where T_i is a compact metric space representing the ambiguous types of player i , and the continuous maps $\gamma_i : T_i \rightarrow \mathcal{K} \Delta(\Theta \times T_{-i})$ associate to each type t_i a compact set of measures $\gamma_i(t_i)$, which can be interpreted as the set of all distributions over Θ and opponents' beliefs that a type t_i views as possible. Of course, non-ambiguous type spaces can be viewed as trivial ambiguous type spaces where the belief sets are singletons. Since we will later define a class of ambiguous type spaces based on a given non-ambiguous type space (S, η) , it is helpful to have a distinct notation for ambiguous type spaces. We will often consider ambiguous type sets that are finite, in which case we can view them as endowed with the discrete topology.

Ahn (2007) shows that the construction of a universal type space from Brandenburger and Dekel (1993) extends in a natural way to a corresponding ambiguous universal type space where types are defined by infinite hierarchies of ambiguous beliefs. The belief hierarchies are derived inductively, letting

$$\begin{aligned} X_i^0 &= \Theta \\ X_i^1 &= X_i^0 \times \prod_{j \neq i} \mathcal{K} \Delta X_j^0 \\ &\vdots \\ X_i^k &= X_i^{k-1} \times \prod_{j \neq i} \mathcal{K} \Delta X_j^{k-1} \\ &\vdots \end{aligned}$$

An ambiguous belief hierarchy for player i is a sequence $b_i = (b_{i1}, b_{i2}, \dots)$, such that for each k , $b_{ik} \in \mathcal{K} \Delta X_i^{k-1}$. Let $\bar{T}_i^0 = \prod_{k=0}^{\infty} \mathcal{K} \Delta X_i^k$, so $b_i \in \bar{T}_i^0$. A sequence b_i is *coherent* if for all $k \geq 1$, $b_{ik} = \text{marg}_{X_i^{k-1}}^{\mathcal{K}} b_{i(k+1)}$. If $\bar{T}_i \subset \bar{T}_i^0$ is defined as the set of ambiguous belief hierarchies that are coherent and for which coherence is common knowledge, then the main result of Ahn (2007) states that there exists a homeomorphism $g_i : \bar{T}_i \rightarrow \mathcal{K} \Delta(\Theta \times \bar{T}_{-i})$. Thus, each element of $\bar{T}_i$ defines a unique compact set of probability measures on $\Theta \times \bar{T}_{-i}$, i.e., the sets $\bar{T}_i$ define an ambiguous type space. Furthermore, Ann shows that this type space is universal, in the sense that any ambiguous type space can be continuously embedded into it, in a manner that preserves the corresponding higher-order beliefs.

The standard approach to embed a type space into the ambiguous universal type space recursively “unpacks” the higher-order beliefs associated to each type. Let (T, γ)

denote any ambiguous type space. As a special case, (T, γ) could represent some non-ambiguous type space (S, η) , in which case $T_i = S_i$ and $\gamma_i(s_i) = \{\eta_i(s_i)\}$ for every i and s_i . To derive the first-order beliefs associated to a type t_i , note that the projection map $\text{proj}_\Theta : \Theta \times T_{-i} \rightarrow \Theta$ is continuous, which implies that $\mathcal{L}_{\text{proj}_\Theta} : \Delta(\Theta \times T_{-i}) \rightarrow \Delta\Theta$, and hence $\mathcal{L}_{\text{proj}_\Theta}^{\mathcal{K}} : \mathcal{K}\Delta(\Theta \times T_{-i}) \rightarrow \mathcal{K}\Delta\Theta$, are continuous. The map $\beta_{i1} : T_i \rightarrow \mathcal{K}\Delta\Theta$ that assigns to a type its first-order beliefs is then defined by $\beta_{i1} := \mathcal{L}_{\text{proj}_\Theta}^{\mathcal{K}} \circ \gamma_i$, and is continuous because it is a composition of two continuous functions. Similarly, to define second-order beliefs, let $(\text{id}, \beta_{-i1}) : \Theta \times T_{-i} \rightarrow \Theta \times \prod_{j \neq i} \mathcal{K}\Delta\Theta$ denote the map that assigns to each tuple (θ, t_{-i}) the tuple consisting of θ and the first-order beliefs of the types defining t_{-i} , as induced by the profile of maps β_{j1} across all $j \neq i$, denoted by β_{-i1} . Then $\mathcal{L}_{(\text{id}, \beta_{-i1})} : \Delta(\Theta \times T_{-i}) \rightarrow \Delta(\Theta \times \prod_{j \neq i} \mathcal{K}\Delta\Theta)$ and $\mathcal{L}_{(\text{id}, \beta_{-i1})}^{\mathcal{K}} : \mathcal{K}\Delta(\Theta \times T_{-i}) \rightarrow \mathcal{K}\Delta(\Theta \times \prod_{j \neq i} \mathcal{K}\Delta\Theta)$, and we can define a map $\beta_{i2} := \mathcal{L}_{(\text{id}, \beta_{-i1})}^{\mathcal{K}} \circ \gamma_i$ that assigns to each type t_i its corresponding (ambiguous) second-order beliefs. Proceeding inductively, we can derive all higher-order beliefs, by letting, $\beta_i^k := (\beta_{i1}, \dots, \beta_{ik})$, and defining $(k+1)$ -th order beliefs by $\beta_{i(k+1)} : T_i \rightarrow \mathcal{K}\Delta X_i^k$ with $\beta_{i(k+1)} := \mathcal{L}_{(\text{id}, \beta_{-i}^k)}^{\mathcal{K}} \circ \gamma_i$. The resulting map $\beta_i := (\beta_{i1}, \beta_{i2}, \dots)$ assigns to each type t_i an ambiguous belief hierarchy in $\prod_{k=0}^{\infty} \mathcal{K}\Delta X_i^k$. Moreover, Proposition 7 in [Ahn \(2007\)](#) implies that the resulting hierarchies are elements of the universal ambiguous type space, and thus that the maps β_i define an embedding of the type space (T, γ) into the ambiguous universal type space $(\bar{T}, g)$. We can therefore view β_i as a map from T_i to $\bar{T}_i$.

Following [Mertens and Zamir \(1985\)](#), a type space (T, γ) is said to satisfy non-redundancy if the maps β_i are injective. In the context of non-ambiguous type spaces, [Liu \(2009\)](#) shows that under very mild conditions, there exists, for any type space that does not satisfy non-redundancy, a corresponding non-redundant type space based on an expanded set of parameters whose belief hierarchies, when appropriately marginalized, coincide with the belief hierarchies of the initial type space. Moreover, the non-redundant type space based on the expanded set of parameters yields the same Bayesian equilibrium predictions as the initial redundant type space. The interpretation given for this result is that redundant type spaces capture “hidden parameters” that the players know but the analyst does not know. The conditions required for the result in [Liu \(2009\)](#) are satisfied in the topological environment we consider, so when necessary, we will assume that the non-ambiguous type spaces in our analysis satisfy non-redundancy.

Consider then our objective of approximating non-ambiguous type spaces with ambiguous type spaces, and follow [Liu \(2009\)](#) and interpret redundant type structures as a result of missing parameters in the specification of the model. If we start with a non-redundant non-ambiguous type space, we can think of a non-redundant ambiguous approximation as arising only from errors in the players’ belief formation process regarding beliefs and higher-order beliefs over the given parameter space, and not from errors in specifying the parameter space. While most of the specific approximating type spaces we construct inherit the non-redundancy of the initial non-ambiguous type space, when restricting our analysis to non-redundant approximations, we will main-

tain the interpretation that this is a result of no additional missing parameters being introduced to the model in the process of adding ambiguity. Hence, the resulting framework is not applicable to analyzing robustness to a misspecification of the parameter space.

3 Perturbing type spaces

Consider a researcher who wants to analyze a game played in an environment with incomplete information. Based on his knowledge of the corresponding environment, this analyst constructs a model that represents the situation at hand. The model specifies the possible payoffs resulting from the game as a function of a parameter space Θ , and the beliefs and higher-order beliefs that the players of the game may hold. Thus, assuming a fixed parameter space and a fixed specification of payoff functions, the analyst's model will be given by some non-ambiguous or ambiguous type space $\mathcal{T}' \equiv (T', \gamma')$ that reflects any restrictions he derived regarding the interacting belief structure associated with the strategic environment. Note that in order to choose a particular model of the game, the analyst must specify all the belief hierarchies that are possibly held by the players according to his knowledge of the game. The analyst's predictions (based on some solution concept he uses) are robust to a misspecification of the belief structure reflected in the model if small errors in the specification of the model have only a minor effect on the resulting predictions. A misspecification of the analyst's model relative to the beliefs and higher-order beliefs entertained by the players can arise from explicit errors in the chosen belief hierarchies, or from choosing a type space defined by sets of belief hierarchies that are in some sense either too large or too small. Let $\mathcal{T} \equiv (T, \gamma)$ denote a type space that represents the actual belief hierarchies of the players. A definition of robustness requires a notion of when the analyst's type space $\mathcal{T}'$ can be considered to be close to the players' type space $\mathcal{T}$. We next introduce a metric on the set of type spaces that provides such a notion.

3.1 A metric on the set of type spaces

Two abstract Θ -based ambiguous or non-ambiguous type spaces $\mathcal{T}' \equiv (T', \gamma')$ and $\mathcal{T} \equiv (T, \gamma)$ are conceptually distinct structures, since the measures defining the corresponding beliefs will in general have distinct support. Defining a measure of distance between the two thus requires that they can both be meaningfully embedded into a common structure, and that this common structure is endowed with a metric or topology that can be interpreted as capturing possible errors in the analyst's modeling process. Clearly, a common structure that appropriately captures the beliefs associated to types is given by the ambiguous universal type space $\tilde{\mathcal{T}} \equiv (\tilde{T}, g)$. Furthermore, any ambiguous or non-ambiguous type space can be embedded into $\tilde{\mathcal{T}}$ using the maps β_i described in the previous section.⁸ The belief hierarchies comprising the ambiguous universal type space can easily be endowed with a metrizable product topology. As

⁸ Note that for non-redundant type spaces, this embedding does not lose any information about the associated belief structures.

long as Θ is compact, all levels of the belief hierarchies can be metrized using a Hausdorff metric derived from the respective Prohorov metric. The product topology on the belief hierarchies can then in turn be metrized using an appropriate product metric, as shown in [Dudley \(2002, pp. 50–51\)](#). Denote the corresponding product metric using $\bar{d}$. While the product topology associated to the metric $\bar{d}$ is known to be a weak topology in the sense that some solution concepts do not yield continuous predictions (see [Dekel et al. 2006](#)), it captures many natural inherent properties of belief formation, as discussed in [Weinstein and Yildiz \(2007, pp. 371–372\)](#). It is thus arguably a topology that can provide an appropriate basis for robustness analyses by capturing potential errors in the analyst's modeling process.

To define a metric over the set of all Θ -based ambiguous type spaces, denote this set by $\mathbf{T}$, and view the set of non-ambiguous type spaces as a subset of $\mathbf{T}$, using the usual interpretation of precise beliefs as singleton sets. We can then define a metric on $\mathbf{T}$ based on the Hausdorff metric applied to the embeddings of the type sets T_i , $\beta_i(T_i)$, into the ambiguous universal type space $\bar{T}$. Let $\bar{d}^H$ denote the Hausdorff metric induced by the metric $\bar{d}$ on $\bar{T}$, and let $\mathcal{T} = (T, \gamma)$ and $\mathcal{T}' = (T', \gamma')$ denote two type spaces in $\mathbf{T}$. A pseudometric on $\mathbf{T}$ is defined by the function $\mathbf{d} : \mathbf{T} \times \mathbf{T} \rightarrow \mathbb{R}_+$, with

$$\mathbf{d}(\mathcal{T}, \mathcal{T}') := \max_i \bar{d}^H(\beta_i(T_i), \beta_i(T'_i)).$$

Since we required type sets T_i to be compact, and belief maps γ_i to be continuous, the embedding β_i is continuous, and hence the images of the type sets T_i and T'_i in $\bar{T}_i$ are compact. The function $\mathbf{d}$ clearly satisfies symmetry and the triangle inequality, so it does yield a pseudometric on $\mathbf{T}$.

$\mathbf{d}$ is only a metric on the set of equivalence classes of type spaces that induce the same belief hierarchies in $\bar{T}$. Type spaces that induce the same hierarchies of beliefs can be related using the concept of a hierarchy morphism, which has been extensively analyzed in the literature (see [Heifetz and Samet 1998](#); [Friedenberg and Meier 2011](#)).

Definition 1 Let $\mathcal{T} = (T, \gamma)$ and $\mathcal{T}' = (T', \gamma')$ be two Θ -based ambiguous type spaces, and denote the corresponding embeddings into $\bar{T}$ by $(\beta_i)_{i \in N}$ and $(\beta'_i)_{i \in N}$, respectively. A *hierarchy morphism* from $\mathcal{T}$ to $\mathcal{T}'$ is defined by a measurable map $h = (h_i)_{i \in N} : T \rightarrow T'$, such that $\beta'_i \circ h_i = \beta_i$ for every i .

Assuming that all type spaces under consideration are non-redundant, the fact that type sets are compact and the belief maps continuous, implies that β_i and β'_i are homeomorphisms between the type sets T_i and T'_i and their images in $\bar{T}_i$. Thus, for every open set $O \subset T'_i$, there exists an open set $U \subset \bar{T}_i$ such that $\beta'_i(O) = U$. Then if h_i is a hierarchy morphism, $h_i^{-1}(O) = h_i^{-1}([\beta'_i]^{-1}(U)) = \beta_i^{-1}(U)$, which implies that $h_i^{-1}(O)$ is open, and thus that h_i is continuous. The non-redundancy of the two type spaces implies that h_i must be injective, and is therefore a homeomorphism between T_i and its image $h_i(T_i)$ in T'_i . It follows that if a hierarchy morphism is surjective, it defines a homeomorphism between T and T' , and h^{-1} is a hierarchy morphism from $\mathcal{T}'$ to $\mathcal{T}$. Note also that the non-redundancy of $\mathcal{T}$ and $\mathcal{T}'$ implies that the embeddings β and β' preserve the beliefs associated to types, which implies that a hierarchy morphism h is also a type morphism, i.e., a map that preserves the

structure of the two spaces as defined by the belief maps. The following definition of a type morphism was first introduced in [Mertens and Zamir \(1985\)](#):

Definition 2 A *type morphism* from $\mathcal{T} = (T, \gamma)$ to $\mathcal{T}' = (T', \gamma')$ is a measurable map $h = (h_i)_{i \in N} : T \rightarrow T'$, such that $\mathcal{L}_{(\text{id}, h_{-i})} \circ \gamma_i = \gamma'_i \circ h_i$ for every i .

In the context of non-ambiguous type spaces, [Heifetz and Samet \(1998, Proposition 5.1\)](#) show that every type morphism is also a hierarchy morphism. Their proof translates directly to the ambiguous type spaces considered here. Together with the remarks made above, we get the following result:

Lemma 1 A map $h = (h_i)_{i \in N} : T \rightarrow T'$ is a hierarchy morphism from $\mathcal{T} = (T, \gamma)$ to $\mathcal{T}' = (T', \gamma')$ if and only if it is a type morphism.

Note that Lemma 1 may not hold if the compactness and non-redundancy assumptions we made are not satisfied.⁹

If a hierarchy morphism, and hence a type morphism, h , is a homeomorphism between $\mathcal{T}$ and $\mathcal{T}'$, it is said to be a *type isomorphism*, and the two spaces are said to be *isomorphic*. Two isomorphic type spaces are essentially identical in terms of their belief structures, and can therefore be considered identical except for a renaming of types. The pseudometric $\mathbf{d}$ captures this notion of sameness:

Proposition 2 Two type spaces $\mathcal{T}$ and $\mathcal{T}'$ are isomorphic if and only if $\mathbf{d}(\mathcal{T}, \mathcal{T}') = 0$.

Proof By definition, two isomorphic type spaces generate the same compact set of belief hierarchies, which yields $\mathbf{d}(\mathcal{T}, \mathcal{T}') = 0$. Conversely, if

$$\mathbf{d}(\mathcal{T}, \mathcal{T}') = \max_i \bar{d}^H(\beta_i(T_i), \beta_i(T'_i)) = 0,$$

the non-redundancy of the type spaces and the compactness of $\beta_i(T_i)$ and $\beta_i(T'_i)$ implies that for each $t_i \in T_i$, there exists a unique $t'_i \in T'_i$ such that

$$\{t'_i\} = \arg \min_{r'_i \in T'_i} \bar{d}(\beta(t_i), \beta(r'_i)),$$

and $\beta_i(t_i) = \beta_i(t'_i)$, so we can define $h_i(t_i) = t'_i$. □

Hence, if we let $\mathbf{T}^*$ denote the set of equivalence classes of isomorphic type spaces, then $\mathbf{d}$ defines a metric on $\mathbf{T}^*$. These considerations show that if we think of type isomorphisms as the correct way to describe when two type spaces can be viewed as identical, then the metric $\mathbf{d}$ can be seen as a generalization of the notion of a type

⁹ [Friedenberg and Meier \(2011\)](#) provide a thorough analysis of the relation between hierarchy morphisms and type morphisms for non-ambiguous type spaces having a more general mathematical structure. Specifically, they show that for type spaces that satisfy the compactness and non-redundancy assumptions made in this paper, hierarchy morphisms and type morphisms are equivalent ([Friedenberg and Meier 2011, Corollary 7.2](#)).

isomorphism, in the sense that it characterizes how similar two non-isomorphic type spaces are.

Returning to the interpretation of $\mathcal{T}'$ as the analyst's model, and $\mathcal{T}$ as the true representation of the players' belief structures, if $\mathbf{d}(\mathcal{T}, \mathcal{T}') \leq \varepsilon$ for some $\varepsilon > 0$, then each belief hierarchy corresponding to the analyst's model is at a distance of at most ε from some belief hierarchy in the players' type structure. In particular, for each type $t_i \in T_i$, the set of types in T'_i whose induced belief hierarchies lie within a (closed) ε -neighborhood of $\beta_i(t_i)$ according to the metric $\bar{d}$, includes all types in T'_i that can be viewed as approximations of t_i . Define the set of all such types by

$$\kappa_i(t_i) := \{t'_i \in T'_i \mid \bar{d}(\beta_i(t_i), \beta_i(t'_i)) \leq \varepsilon\}.$$

Then $\bar{d}^H(\{\beta_i(t_i)\}, \beta_i(\kappa_i(t_i))) \leq \varepsilon$ and $\bigcup_{t_i \in T_i} \kappa_i(t_i) = T'_i$, which implies that T'_i only contains types that approximate some type in the players' type structure. Hence, the analyst's model $\mathcal{T}'$ yields an approximation of the players' type structure in the sense that the belief hierarchy of every type in $\mathcal{T}'$ is close to a belief hierarchy contained in the players' type structure. This implies that $\beta_i(T'_i) \subseteq \bar{\mathcal{N}}_\varepsilon(\beta_i(T_i))$, where $\bar{\mathcal{N}}_\varepsilon$ denotes the closed ε -neighborhood of the respective set. Similarly, it follows that $\beta_i(T_i) \subseteq \bar{\mathcal{N}}_\varepsilon(\beta_i(T'_i))$, so every type in $\mathcal{T}$ is approximated by some type in $\mathcal{T}'$. Hence, closeness of two type spaces in the metric $\mathbf{d}$ yields both closeness of beliefs, in the sense of capturing errors in specifying individual belief hierarchies, and closeness of type spaces, in the sense of capturing errors in specifying a type space as the context in which the game is played.

Conversely, an analogous interpretation implies that if the analyst is convinced that he did not make any significant errors in his modeling process, he should expect the true type structure $\mathcal{T}$ to be at a distance of at most ε from his model $\mathcal{T}'$, according to the metric $\mathbf{d}$. We can then view any type space $\mathcal{T}$ with $\mathbf{d}(\mathcal{T}, \mathcal{T}') \leq \varepsilon$ as a perturbation of the analyst's model $\mathcal{T}'$. A general approach to robustness would then consider a solution concept Γ that generates a prediction for any type space associated to a given game, and would require that for every $\delta > 0$, there exists some $\varepsilon > 0$ so that for a class of perturbations $\mathcal{T}$ with $\mathbf{d}(\mathcal{T}, \mathcal{T}') \leq \varepsilon$, the predictions $\Gamma(\mathcal{T})$ are within a δ -neighborhood of $\Gamma(\mathcal{T}')$, measured appropriately. This approach to defining robustness is analogous to the usual definition of qualitative robustness in statistical settings, as described in Huber (1981, pp. 7–10).

3.2 Robustness without ambiguity

While our main objective is to provide foundations for analyzing robustness of solution concepts to ambiguity of beliefs, the metric $\mathbf{d}$, suitably restricted to the set of non-ambiguous type spaces, clearly also yields a notion of distance that can be used to analyze robustness to misspecification of non-ambiguous belief structures, when ambiguous beliefs are not considered. We next compare the resulting approach to robustness to other approaches found in the literature.

The literature has so far analyzed robustness for two distinct classes of situations with incomplete (but not ambiguous) beliefs. The first class considers games that

have an ex-ante stage where a move of nature determines the players' types according to a prior distribution. The belief structure in such games is completely determined by the specification of the prior, and a perturbation can be defined by an alternative prior that is in some way close to the initial prior. The robustness literature using such an *ex-ante approach* was initiated by [Kajii and Morris \(1997\)](#), who considered whether equilibrium predictions for static games with complete information are robust to the introduction of a small amount of incomplete information. In their model, the belief structure associated to the initial model of the analyst is defined by a complete information game, and the corresponding incomplete information perturbations are derived from priors of incomplete information games that place high probability on states of the world where all players' payoffs are equal to the payoffs of the initial complete information game. The definition of robustness requires the ex-ante equilibrium action distribution resulting from the incomplete information perturbations to be close to the equilibrium action distribution of the complete information game.¹⁰ Since perturbations in the ex-ante approach are only based on the probability of playing the initial complete information game, and do not involve any assumptions on higher-order beliefs, the resulting belief structures need not be close to the initial belief structure according to our metric $\mathbf{d}$.

The second class of robustness models considers situations without an ex-ante stage, where the primitives of the model are the players' belief hierarchies, which are usually specified by some type space. This approach, which we follow in this paper, is commonly referred to as an *interim approach*. Papers based on this approach also often use the product topology on belief hierarchies as a measure for the closeness of such hierarchies. A classic result that is related to, but not necessarily formulated as a robustness result, is the result of [Mertens and Zamir \(1985\)](#) that every type space can be approximated by a finite type space, in the sense that the sets of belief hierarchies corresponding to the two type spaces are close in the Hausdorff topology on closed subsets of the universal type space. While the formulation of [Mertens and Zamir \(1985\)](#) is based on a topological framework, restating their result in a metric framework would be equivalent to the use of the metric $\mathbf{d}$. This approximation result has provided the main justification for many papers on Bayesian games to restrict their analysis to finite type spaces.

A major recent contribution to the robustness literature based on the interim approach is the paper by [Weinstein and Yildiz \(2007\)](#). It considers an environment where no restrictions are possible on the set of belief hierarchies, so the belief structure of an incomplete information game can only be represented by the corresponding universal type space. Furthermore, it assumes that action spaces are finite, and that each action of every player is strictly dominant for some element of the parameter space Θ . [Weinstein and Yildiz \(2007\)](#) then use the product topology on belief hierarchies to show that the set of types with a unique interim correlated rationalizable (ICR) action is open and dense in the universal type space.¹¹ The main difference between our approach and

¹⁰ Extensions of [Kajii and Morris \(1997\)](#) are explored by [Haimanko and Kajii \(2012\)](#), who only require approximate ε -equilibria for the perturbed games, [Oyama and Tercieux \(2010\)](#), who allow for heterogeneous priors, and [Chassang and Takahashi \(2011\)](#), who analyze repeated games.

¹¹ [Penta \(2012\)](#) extends these results to dynamic games.

the setup of [Weinstein and Yildiz \(2007\)](#) is that we allow the analyst and the players to exclude some belief hierarchies from consideration when analyzing the game. Note that the set of ICR actions of a type is type space invariant, in the sense that it is determined as a function of the corresponding belief hierarchy independently of the type space under consideration (see [Dekel et al. 2007](#); [Penta 2012](#)). However, analogous invariance properties do not hold for all solution concepts, as discussed in the following paragraph. Thus, if we allow an analyst to restrict the set of belief hierarchies under consideration based on his understanding of the strategic situation, a robustness analysis that only considers an embedding of types into the universal type space may exclude certain outcomes from being robust just because the universal type space is too large relative to the analyst's initial type structure. In such a situation, considering perturbations that are close according to $\mathbf{d}$ may be more appropriate to determine robustness. Notice also that even if we are only interested in analyzing robustness with respect to non-ambiguous perturbations, the results of [Weinstein and Yildiz \(2007\)](#) still leave room for a variety of interesting questions. For example, one could consider games with a continuum of actions, or devise a robustness test where the equilibrium actions in the initial model are only required to induce ε -best responses in nearby perturbations (see, for example, [Stauber 2011](#); [Haimanko and Kajii 2012](#) for such approaches).

An example of a solution concept that is not type space invariant is discussed by [Friedenberg and Meier \(2012\)](#), who show that there may exist Bayesian Nash equilibria for a smaller type space $\mathcal{T}$ that cannot be extended to a larger type space $\mathcal{T}'$ that contains the smaller space $\mathcal{T}$. The interpretation given is that $\mathcal{T}'$ is the analyst's type space, and that the players know more about the strategic situation than the analyst and can therefore exclude more belief hierarchies from consideration, thus basing their actions on a smaller type space $\mathcal{T}$. The extension failure identified by [Friedenberg and Meier \(2012\)](#) can be interpreted as a lack of robustness to a misspecification of the type space by the analyst, in the sense that the analyst's prediction can miss outcomes that are feasible under the players' belief structure. Our approach to robustness only allows for small differences in the analyst's and the players' type structures that can plausibly arise from small errors in the analyst's modeling process. Thus, it does not consider as perturbations type spaces that are significantly smaller or significantly larger than the analyst's model. If the solution concept used by the analyst is not type space invariant, a perturbation that is too small, or too large (such as the universal type space), could yield a failure of robustness that does not result from small errors in the analyst's model.

The use of the product topology to characterize nearby beliefs is based on an interpretation of this topology as capturing natural properties of belief formation and associated errors. Starting from this interpretation, a robustness analysis is intended to identify solution concepts which do not lose their optimality properties for nearby beliefs. An alternative approach to defining closeness of beliefs within the interim setup is taken in [Dekel et al. \(2006\)](#), who consider a given solution concept, in their case ε correlated interim rationalizability, and introduce a topology on types in the universal type space under which this solution concept is continuous. This topology can be viewed as providing conditions on the belief formation process that guarantee that a given solution concept is robust, as long as errors in beliefs can be modeled using the given topology. In this sense, this approach and the approach based on the product topology are orthogonal—one fixes a way to model errors in beliefs and pro-

vides restrictions on solution concepts, the other fixes a solution concept and provides restrictions on modeling errors in beliefs.

3.3 Ambiguous perturbations

The metric $\mathbf{d}$ describes when a type space can be viewed as approximating another type space, independently of whether any of the two type spaces are ambiguous or not.¹² Henceforth, we will consider a given Θ -based non-ambiguous type space $\mathcal{S} \equiv (S, \eta)$, which we interpret as the analyst's model, and characterize methods that can be used to construct ambiguous perturbations of $\mathcal{S}$ that are consistent with the metric $\mathbf{d}$. The resulting methods provide a foundation for analyzing whether the use of a non-ambiguous model by the analyst is robust in an environment where players may not have enough information to form precise beliefs. Letting $\gamma_i(s_i) := \{\eta_i(s_i)\}$, we can use the maps β_i described in the previous section to define an embedding $\beta_i : S_i \rightarrow \bar{T}_i$ into the ambiguous universal type space, and interpret $\mathcal{S}$ as an ambiguous type space with singleton belief sets. Even though we can view every ambiguous Θ -based type space $\mathcal{T} \equiv (T, \gamma)$ that satisfies $\mathbf{d}(\mathcal{S}, \mathcal{T}) \leq \varepsilon$ for some small $\varepsilon > 0$ as a perturbation of $\mathcal{S}$, working with all such perturbations is often impractical from an applied viewpoint. Moreover, the characterization using $\mathbf{d}$ provides no tractable procedure to construct such perturbations, or to characterize their convergence to the unperturbed belief structure. We now introduce an alternative way to define a sub-class of such ambiguous perturbations, which we use in the following sections to derive two distinct methods to construct perturbations that allow a simple characterization of convergence to the initial type space $\mathcal{S}$.

Assume that $\mathbf{d}(\mathcal{S}, \mathcal{T}) \leq \varepsilon$ for some ambiguous type space $\mathcal{T} \equiv (T, \gamma)$, and define correspondences $\tilde{\alpha}_i : S_i \rightrightarrows T_i$ by

$$\tilde{\alpha}_i(s_i) := \{t_i \in T_i \mid \bar{d}(\beta_i(s_i), \beta_i(t_i)) \leq \varepsilon\}.$$

Then $\bar{d}^H(\{\beta_i(s_i)\}, \beta_i(\tilde{\alpha}_i(s_i))) \leq \varepsilon$, so every ambiguous type in $\tilde{\alpha}_i(s_i)$ can be viewed as approximating the type s_i , and $\bigcup_{s_i \in S_i} \tilde{\alpha}_i(s_i) = T_i$. Conversely, given type spaces $\mathcal{S}$ and $\mathcal{T}$ and a collection of correspondences $\tilde{\alpha}_i : S_i \rightrightarrows T_i$, such that

$$\bar{d}^H(\{\beta_i(s_i)\}, \beta_i(\tilde{\alpha}_i(s_i))) \leq \varepsilon \quad \text{and} \quad \bigcup_{s_i \in S_i} \tilde{\alpha}_i(s_i) = T_i,$$

it must be the case that $\mathbf{d}(\mathcal{S}, \mathcal{T}) \leq \varepsilon$. Since intuitively, an ambiguous perturbation should be expected to yield a belief structure that is coarser than the initial belief structure, we will restrict our analysis to a class of perturbations that are defined by such correspondences $\tilde{\alpha}_i$, but where in addition the correspondences take on values that only contain a single element, so they are in fact functions:

¹² To our knowledge, except for the result of [Mertens and Zamir \(1985\)](#) that any (non-ambiguous) type space can be approximated by a finite type space, this metric has not been used in the literature to explicitly analyze robustness questions.

Definition 3 A Θ -based ambiguous type space $\mathcal{T} = (T, \gamma)$ is an *ambiguous ε -perturbation* of the non-ambiguous type space $\mathcal{S} = (S, \eta)$ for some $\varepsilon > 0$, if for each $i \in N$, there exists a measurable surjective map $\alpha_i : S_i \rightarrow T_i$, such that

$$\bar{d}(\beta_i(s_i), \beta_i(\alpha_i(s_i))) \leq \varepsilon \quad \text{for every } s_i \in S_i.$$

Thus, any ambiguous ε -perturbation $\mathcal{T}$ of $\mathcal{S}$ satisfies $\mathbf{d}(\mathcal{S}, \mathcal{T}) \leq \varepsilon$, but the converse need not be true. If in addition,

$$\varepsilon = \max_i \sup_{s_i \in S_i} \bar{d}(\beta_i(s_i), \beta_i(\alpha_i(s_i))),$$

we will refer to ε as the *precision* of the corresponding perturbation. At the risk of confusing the reader, we will say that the precision increases as ε decreases. The requirement that the maps α_i are surjective implies that $\alpha_i^{-1}(\{t_i\})$ is non-empty for each $t_i \in T_i$, so the definition requires that every type t_i in the perturbed type space (T, γ) can be interpreted as approximating the beliefs of at least one type s_i from the initial type space.¹³ Although the definition does not explicitly require that the type space (T, γ) contains only purely ambiguous types for which beliefs are not singletons, we would expect the most interesting or useful perturbations to be of this type. Simple procedures to construct such perturbations are introduced in the following two sections.

An ambiguous ε -perturbation (T, γ) is defined to be *finite* if the corresponding type sets T_i are finite sets. Given a finite ε -perturbation (T, γ) , the surjectivity of the maps α_i implies that for each i , the sets $\alpha_i^{-1}(\{t_i\})$ defined by all $t_i \in T_i$ form a finite partition of the non-ambiguous type set S_i . Denote the corresponding partition by $\pi_i^{\alpha_i}$. For a given t_i and any two $s_i, s'_i \in \alpha_i^{-1}(\{t_i\})$, the triangle inequality for the metric $\bar{d}$ together with the definition of an ε -perturbation imply that $\bar{d}(\beta_i(s_i), \beta_i(s'_i)) \leq 2\varepsilon$. Since β_i is a continuous embedding of the compact set S_i into $\bar{T}_i$ (Ahn 2007, Prop. 7), if S_i is assumed to be non-redundant, β_i is injective, and it follows that β_i is a homeomorphism between S_i and its image $\beta_i(S_i)$ in $\bar{T}_i$ (see Dudley 2002 p. 40). Thus β_i and its inverse are continuous, and hence uniformly continuous, since the image of S_i under β_i is compact. This implies that if the precision ε in a sequence of ε -perturbations approaches zero, then the mesh of the partitions on S_i defined by these perturbations must also converge to zero. We summarize this observation in the following proposition:

Proposition 3 Let (T^n, γ^n) be a sequence of finite ambiguous perturbations of a non-redundant type space (S, η) , with corresponding precision levels ε^n and type maps α_i^n , and assume that $\varepsilon^n \rightarrow 0$ as $n \rightarrow \infty$. Then the mesh of the partitions $\pi_i^{\alpha_i^n}$ converges to zero in n , i.e., $\mathcal{M}(\pi_i^{\alpha_i^n}) \rightarrow 0$ as $n \rightarrow \infty$.

The behavior of such sequences of perturbations as their precision converges to zero is important for robustness, as the usual qualitative definitions of robustness can

¹³ A similar definition of a perturbation, which however does not require surjectivity, is employed in Weinstein and Yildiz (2007, p. 386) to introduce a robustness notion in a setting without ambiguity.

be framed in terms of perturbations that converge to the unperturbed environment.¹⁴ Accordingly, given an initial type space (S, η) and a sequence of ambiguous type spaces (T^n, γ^n) as considered in Proposition 3, a solution concept Γ is deemed to be robust if $\Gamma(T^n, \gamma^n) \rightarrow \Gamma(S, \eta)$ (using some appropriate notion of convergence) as $\varepsilon^n \rightarrow 0$.

4 Partitional type spaces

In the previous section, we showed that any finite ambiguous perturbation defines a finite partition of the initial type space, whose mesh will be decreasing as the precision of the perturbation increases. In this section, we prove a partial converse of this result. We show that any finite partition of an initial non-ambiguous type space can be viewed as defining a corresponding finite ambiguous perturbation, and show that as the mesh of the defining partition decreases, the precision of the resulting perturbation will increase. The resulting construction yields a simple procedure to derive a class of finite perturbations and characterize their convergence to the initial unperturbed belief structure. An extension of this construction is shown in the next section to yield the set of all finite perturbations.

Given a Θ -based non-ambiguous type space (S, η) , consider a sequence $\pi^n = (\pi_i^n)_{i \in N}$ of measurable finite partitions of S_i , with $\pi_i^n(s_i)$ denoting the partition element containing the type s_i . Assume that $\mathcal{M}(\pi^n) \rightarrow 0$, i.e., the mesh of the partitions converges to zero in n .¹⁵ The existence of such a sequence of finite partitions follows from the compactness of the type sets S_i (see Stauber 2011, Lemma 1). To simplify the notation, we will write $\pi^n \rightarrow 0$ whenever $\mathcal{M}(\pi^n) \rightarrow 0$.

For each partition π^n , we now define an associated *partitional ambiguous type space* (T^n, γ^n) .¹⁶ Letting

$$T_i^n := \{\pi_i^n(s_i) \mid s_i \in S_i\},$$

we identify each partition element of π_i^n with a corresponding ambiguous type. The finiteness of the partitions implies that the type sets T_i^n are finite, and hence compact metrizable. To define the belief maps γ_i^n , let $t_i^n : S_i \rightarrow T_i^n$, such that $t_i^n(s_i) := \pi_i^n(s_i)$, so $t_i^n(s_i)$ is just the type in T_i^n derived from the partition element $\pi_i^n(s_i)$.¹⁷ Since each π_i^n is measurable, the image measure of every $\eta_i(s_i)$ under (id, t_{-i}^n) , $\mathcal{L}_{(\text{id}, t_{-i}^n)}(\eta_i(s_i))$, is an element of $\Delta(\Theta \times T_{-i}^n)$. Letting $\eta_i^n := \mathcal{L}_{(\text{id}, t_{-i}^n)} \circ \eta_i$, γ_i^n is defined by

$$\gamma_i^n(t_i^n(s_i)) := \text{cl} \left\{ \eta_i^n(s'_i) \mid s'_i \in \pi_i^n(s_i) \right\}.$$

¹⁴ In this context, finiteness can be interpreted as arising from an intrinsic limitation in the players' ability to form a wide variety of precise or ambiguous beliefs.

¹⁵ Clearly, such partitions will only be of interest if the initial type space includes a continuum of types.

¹⁶ This definition of a partitional type space was introduced in Stauber (2011).

¹⁷ Although the maps t_i^n are essentially identical to π_i^n , it is sometimes useful to think of $t_i^n(s_i)$ as an element of T_i^n , and of $\pi_i^n(s_i)$ as a subset of S_i .

Continuity of the maps γ_i^n follows from the finiteness of T_i^n , so each (T^n, γ^n) defines an ambiguous type space. We can interpret the type space (T^n, γ^n) as arising from a situation where the players cannot unambiguously determine beliefs as associated with a “true” type s_i , but are able to restrict their beliefs to those distributions that correspond to some element of $\pi_i^n(s_i)$. If every player’s belief formation is constrained in such a way, the resulting belief structure will be characterized by the type space (T^n, γ^n) .

The partitional type spaces (T^n, γ^n) can be embedded into the ambiguous universal type space $(\bar{T}, g)$ using the maps β_{ik} and β_i defined previously. For our analysis, it will also be useful to define maps $\tilde{\beta}_{ik}^n := \mathcal{L}_{(\text{id}, \beta_{-i}^{k-1})} \circ \eta_i^n$, which assign to each non-ambiguous type s_i the higher-order beliefs induced by this type’s beliefs about ambiguous types in T_{-i}^n , as given by $\eta_i^n(s_i)$. It follows that

$$\beta_{ik}(t_i^n(s_i)) = \text{cl}\{\tilde{\beta}_{ik}^n(s'_i) \mid s'_i \in \pi_i^n(s_i)\}.$$

We summarize below the various maps used in the subsequent analysis:

$$\begin{aligned} \pi_i^n &: S_i \rightarrow 2^{S_i}, \\ t_i^n &: S_i \rightarrow T_i^n, \\ \eta_i &: S_i \rightarrow \Delta(\Theta \times S_{-i}), \\ \eta_i^n &: S_i \rightarrow \Delta(\Theta \times T_{-i}^n), \\ \gamma_i^n &: T_i^n \rightarrow \mathcal{K} \Delta(\Theta \times T_{-i}^n), \\ \beta_{ik} &: S_i \text{ (or) } T_i^n \rightarrow \mathcal{K} \Delta X_i^{k-1}, \\ \tilde{\beta}_{ik}^n &: S_i \rightarrow \Delta X_i^{k-1}. \end{aligned}$$

Clearly, the partitional type spaces (T^n, γ^n) can be viewed as finite ambiguous perturbations of the type space (S, η) , as defined in the previous section. The corresponding type maps are given by $\alpha_i^n := t_i^n$, and the precision levels by

$$\varepsilon^n := \max_i \sup_{s_i} \bar{d}(\beta_i(s_i), \beta_i(t_i^n(s_i))).$$

This approach to deriving ambiguous perturbations is very tractable, but given the interpretation of types as an implicit representation of beliefs and belief hierarchies, its implication regarding the resulting ambiguous belief hierarchies, and hence the resulting precision of the perturbations, is unclear. Intuitively, we would expect the precision of the approximation to increase as the mesh of the partitions decreases. We now prove that this is indeed the case. Using an inductive argument to show that each level of the corresponding belief hierarchies converges in the respective Hausdorff metric, we show that the ambiguous types belonging to partitional type spaces approximate the original non-ambiguous types, in the sense that the ambiguous belief hierarchy induced by a type $t_i^n(s_i)$, $\beta_i(t_i^n(s_i))$, converges in the product topology induced by the weak topology on individual belief levels to the belief hierarchy of the non-ambiguous type s_i , $\beta_i(s_i)$, as the mesh of the partitions converges to zero. Furthermore, the precision of the ambiguous perturbations defined by the respective partitions converges to zero as their mesh converges to zero.

To see that first-order beliefs converge, note that since the projection on Θ is continuous, $\mathcal{L}_{\text{proj}_\Theta}$ is continuous, and hence $\tilde{\beta}_{i1} := \mathcal{L}_{\text{proj}_\Theta} \circ \eta_i$ is a continuous map from S_i to $\Delta\Theta$. In fact, the compactness of S_i implies that $\tilde{\beta}_{i1}$ is uniformly continuous. Therefore, if $\pi_i^n \rightarrow 0$, we have $\sup_{s_i} \sup_{s'_i \in \pi_i^n(s_i)} d^P(\tilde{\beta}_{i1}(s_i), \tilde{\beta}_{i1}(s'_i)) \rightarrow 0$. Since the marginals on Θ of $\eta_i(s'_i)$ and $\eta_i^n(s'_i)$ are identical,

$$\beta_{i1}(t_i^n(s_i)) = \text{cl}\{\tilde{\beta}_{i1}(s'_i) \mid s'_i \in \pi_i^n(s_i)\},$$

which implies that $\beta_{i1}(t_i^n(s_i)) \rightarrow \beta_{i1}(s_i)$ uniformly on S_i using the appropriate Hausdorff metric, as $\pi_i^n \rightarrow 0$.

Note the difference between $\tilde{\beta}_{i1}(s_i)$ and $\beta_{i1}(s_i)$ —the first yields an element of $\Delta\Theta$, whereas the second yields a singleton set in $\mathcal{K}\Delta\Theta$. Similarly, for $k > 1$, define $\tilde{\beta}_{ik} := \mathcal{L}_{(\text{id}, \beta_{-i}^{k-1})} \circ \eta_i$, so that $\tilde{\beta}_{ik}(s_i)$ again represents the unique element of the set $\beta_{ik}(s_i)$. To complete the convergence proof, we will use an inductive step to show that for each $k > 1$, $\beta_{ik}(t_i^n(s_i))$ converges to $\beta_{ik}(s_i)$ uniformly on S_i in the corresponding Hausdorff metric. The main proof relies on the following lemma:

Lemma 4 *If $\beta_{il}(t_i^n(s_i))$ converges to $\beta_{il}(s_i)$ uniformly on S_i for all i and $l \in \{1, \dots, k-1\}$ as $\pi^n \rightarrow 0$, then $\tilde{\beta}_{ik}^n$ converges to $\tilde{\beta}_{ik}$ uniformly on S_i .*

Proof $\tilde{\beta}_{ik}^n(s_i)$ converges to $\tilde{\beta}_{ik}(s_i)$ in the weak topology, and hence the Prohorov metric, if

$$\lim_n \int f d\tilde{\beta}_{ik}^n(s_i) = \int f d\tilde{\beta}_{ik}(s_i)$$

for every bounded real-valued uniformly continuous function f defined on

$$X_i^{k-1} = \Theta \times \prod_{j \neq i} \mathcal{K} \Delta X_j^0 \times \dots \times \prod_{j \neq i} \mathcal{K} \Delta X_j^{k-2}.$$

Since

$$\tilde{\beta}_{ik}^n = \mathcal{L}_{(\text{id}, \beta_{-i}^{k-1})} \circ \eta_i^n = \mathcal{L}_{(\text{id}, \beta_{-i}^{k-1})} \circ \mathcal{L}_{(\text{id}, t_{-i}^n)} \circ \eta_i = \mathcal{L}_{(\text{id}, \beta_{-i}^{k-1} \circ t_{-i}^n)} \circ \eta_i,$$

and $\tilde{\beta}_{ik} := \mathcal{L}_{(\text{id}, \beta_{-i}^{k-1})} \circ \eta_i$, we have

$$\begin{aligned} & \left| \int_{X_i^{k-1}} f d\tilde{\beta}_{ik}^n(s_i) - \int_{X_i^{k-1}} f d\tilde{\beta}_{ik}(s_i) \right| \\ &= \left| \int_{\Theta \times S_{-i}} [f(\theta, \beta_{-i}^{k-1}(t_{-i}^n(s_{-i}))) - f(\theta, \beta_{-i}^{k-1}(s_{-i}))] d\eta_i(s_i) \right| \\ &\leq \int_{\Theta \times S_{-i}} |f(\theta, \beta_{-i}^{k-1}(t_{-i}^n(s_{-i}))) - f(\theta, \beta_{-i}^{k-1}(s_{-i}))| d\eta_i(s_i) \end{aligned}$$

$$\leq \sup_{(\theta, s_{-i})} \left| f(\theta, \beta_{-i}^{k-1}(t_{-i}^n(s_{-i}))) - f(\theta, \beta_{-i}^{k-1}(s_{-i})) \right|.$$

By assumption, $\beta_{-i}^{k-1}(t_{-i}^n(s_{-i}))$ converges to $\beta_{-i}^{k-1}(s_{-i})$ uniformly on S_{-i} , and f is uniformly continuous, which implies that the last term of the previous equation converges to zero in n . In addition, the converging term is independent of s_i , so $\tilde{\beta}_{ik}^n$ converges to $\tilde{\beta}_{ik}$ uniformly on S_i . $\square$

We can now prove our main result:

Proposition 5 *If $\pi^n \rightarrow 0$, then for each $k \geq 1$, $\beta_{ik}(t_i^n(s_i))$ converges to $\beta_{ik}(s_i)$ uniformly on S_i in the Hausdorff metric on $\mathcal{H} \Delta X_i^{k-1}$.*

Proof We have shown above that $\beta_{i1}(t_i^n(s_i))$ converges to $\beta_{i1}(s_i)$ uniformly on S_i . Assume that $\beta_{il}(t_i^n(s_i))$ converges to $\beta_{il}(s_i)$ uniformly for $l \in \{1, \dots, k-1\}$, and consider the k -th order beliefs, $\beta_{ik}(t_i^n(s_i))$ and $\beta_{ik}(s_i)$.

Let d^H denote the Hausdorff metric on $\mathcal{H} \Delta X_i^{k-1}$ induced by the Prohorov metric d^P , and define

$$\mathcal{H}_{ik}^n(s_i) := \text{cl} \left\{ \tilde{\beta}_{ik}(s'_i) \mid s'_i \in \pi_i^n(s_i) \right\}.$$

Given that $\beta_{ik}(t_i^n(s_i)) = \text{cl} \{ \tilde{\beta}_{ik}^n(s'_i) \mid s'_i \in \pi_i^n(s_i) \}$, Lemma 4 implies that

$$d^H(\beta_{ik}(t_i^n(s_i)), \mathcal{H}_{ik}^n(s_i)) \leq \sup_{s'_i \in \pi_i^n(s_i)} d^P(\tilde{\beta}_{ik}^n(s'_i), \tilde{\beta}_{ik}(s'_i)) \rightarrow 0.$$

Since the convergence shown in Lemma 4 is uniform, it follows that $\beta_{ik}(t_i^n(s_i))$ converges to $\mathcal{H}_{ik}^n(s_i)$ uniformly on S_i . The recursive definition of β_{ik} implies that β_{ik} , and hence $\tilde{\beta}_{ik}$, is uniformly continuous on S_i , which implies that $\mathcal{H}_{ik}^n(s_i)$ converges to $\beta_{ik}(s_i)$ uniformly on S_i . Applying the triangle inequality, we can conclude that $\beta_{ik}(t_i^n(s_i))$ converges uniformly to $\beta_{ik}(s_i)$. $\square$

The fact that the precision levels of the partitional type spaces defined by a sequence of finite partitions π^n will increase as the mesh of the partitions decreases, is now a direct consequence of Proposition 5:

Corollary 6 *Let π^n be a sequence of finite partitions of (S, η) such that $\pi^n \rightarrow 0$, and let (T^n, γ^n) be the partitional ambiguous type spaces defined by π^n . Then the precision levels ε^n associated with (T^n, γ^n) converge to zero in n .*

The convergence characterized by Corollary 6 is equivalent to convergence of (T^n, γ^n) to (S, η) in the metric $\mathbf{d}$:

Proposition 7 *Let π^n be a sequence of finite partitions of a non-ambiguous type space $\mathcal{S} = (S, \eta)$ such that $\pi^n \rightarrow 0$, and let $\mathcal{T}^n = (T^n, \gamma^n)$ be the partitional type spaces defined by π^n . Then $\mathbf{d}(\mathcal{S}, \mathcal{T}^n) \rightarrow 0$ as $n \rightarrow \infty$.*

Proof We need to show that $\bar{d}^H(\beta_i(S_i), \beta_i(T_i^n)) \rightarrow 0$ for every $i \in N$. By Theorem 0.7 in Nadler (1978), this is equivalent to

$$\beta_i(S_i) = \liminf_n \beta_i(T_i^n) = \limsup_n \beta_i(T_i^n),$$

where $\liminf_n \beta_i(T_i^n)$ is the set of limit points of sequences $\beta_i(t_i^n) \in \beta_i(T_i^n)$, and $\limsup_n \beta_i(T_i^n)$ is the set of cluster points of sequences $\beta_i(t_i^n) \in \beta_i(T_i^n)$.¹⁸ By Proposition 5, $\beta_i(s_i) = \lim_n \beta_i(t_i^n(s_i))$ for every $s_i \in S_i$, which implies that $\beta_i(S_i)$ is a subset of $\liminf_n \beta_i(T_i^n)$. Let $\mathcal{N}_\varepsilon(\beta_i(S_i))$ denote an open ε -neighborhood of $\beta_i(S_i)$ with respect to $\bar{d}$. Since each $t_i^n \in T_i^n$ can be written as $t_i^n = t_i^n(s_i)$ for some s_i , by Proposition 5, there exists a $\bar{n}$ such that $\beta_i(T_i^n) \subseteq \mathcal{N}_\varepsilon(\beta_i(S_i))$ for all $n > \bar{n}$. Then it follows from Theorem 1.1.4 in Aubin and Frankowska (1990) that $\limsup_n \beta_i(T_i^n) \subseteq \text{cl}(\beta_i(S_i)) = \beta_i(S_i)$. Since $\liminf_n \beta_i(T_i^n) \subseteq \limsup_n \beta_i(T_i^n)$, this completes the proof. $\square$

5 A general class of perturbations

The construction of a partitional perturbation assigns to each type $s_i \in S_i$ a corresponding partition element $\pi_i(s_i)$, and then uses the image measures under (id, π_{-i}) of all $\eta_i(s'_i)$ corresponding to a type $s'_i \in \pi_i(s_i)$ to derive a set of measures over Θ and ambiguous types. When the initial non-ambiguous type space (S, η) is a suitably large space with a continuum of types, this approach yields a rich class of finite perturbations that also has an intuitive interpretation based on the standard partitional model of knowledge. However, if the initial space (S, η) is finite or countable, partitional perturbations with small enough mesh would just recover (S, η) . This section introduces a wider class of perturbations whose construction builds on the approach used for partitional perturbations, which also yields a tractable approach to derive non-trivial ambiguous perturbations for finite or countable type spaces. Furthermore, this approach allows us to characterize all finite ambiguous perturbations of any non-ambiguous type space.

In the construction of a partitional type space, a partition π_i can be viewed as assigning to each type s_i a set $\eta_i(\pi_i(s_i))$ of measures over $\Theta \times S_{-i}$. While $\pi_i(s_i)$ can be interpreted as an ambiguous type, we need the image measures under (id, π_{-i}) of all elements of $\eta_i(\pi_i(s_i))$ to define the corresponding ambiguous beliefs. To generalize this construction, we start with maps $\tau_i : S_i \rightarrow 2^{\Delta(\Theta \times S_{-i})}$, which assign to each non-ambiguous type s_i a set $\tau_i(s_i)$ of measures over $\Theta \times S_{-i}$. Note that τ_i can be an arbitrary map, so $\tau_i(s_i)$ need not correspond to measures arising from a partition element containing s_i . We will assume throughout our discussion that each τ_i is either compact-valued and continuous with respect to the Hausdorff metric on $\mathcal{H} \Delta(\Theta \times S_{-i})$, or that τ_i has finite range in $2^{\Delta(\Theta \times S_{-i})}$ and is measurable. If $\tau_i(s_i)$ is interpreted as an ambiguous type associated to s_i , its corresponding beliefs, denoted by $\gamma_i^\tau(\tau_i(s_i))$,

¹⁸ See Aubin and Frankowska (1990, Proposition 1.1.2) and Nadler (1978, Definition 0.5).

can be defined using appropriate image measures by letting

$$\gamma_i^\tau(\tau_i(s_i)) := \text{cl}\{\mathcal{L}_{(\text{id}, \tau_{-i})}(p_i) \mid p_i \in \tau_i(s_i)\}.$$

If τ_i is compact-valued and continuous, the belief maps γ_i^τ are continuous since $\gamma_i^\tau \equiv \mathcal{L}_{(\text{id}, \tau_{-i})}^\mathcal{H}$; if τ_i has finite range, the maps γ_i^τ are continuous because they have finite domain. Hence, if we let T_i^τ denote the range of each τ_i , then $(T^\tau, \gamma^\tau) \equiv (T_i^\tau, \gamma_i^\tau)_{i \in N}$ defines an ambiguous type space.

Partitional type spaces clearly provide a specific example of ambiguous type spaces of the class defined above. Alternative examples can be constructed by letting $\tau_i(s_i)$ be a closed ε -neighborhood of $\eta_i(s_i)$ with respect to the Prohorov metric on $\Delta(\Theta \times S_{-i})$, or an ε -contamination of $\eta_i(s_i)$ with respect to some set $\Lambda \subseteq \Delta(\Theta \times S_{-i})$. These two examples yield type spaces where the τ_i 's are compact-valued and continuous, and can be viewed as generalizations of ε -neighborhoods and ε -contaminations, as defined previously for an environment with one-sided incomplete information. Thus, our definition of perturbations derived from maps τ_i clarifies the use of ε -contaminations in the construction of ambiguous approximating type spaces, as employed in [Bose et al. \(2006\)](#) and [Bose and Daripa \(2009\)](#), and places their construction into a broader context of ambiguous perturbations. In particular, [Bose et al. \(2006\)](#) and [Bose and Daripa \(2009\)](#) start with a standard non-ambiguous auction environment where players' types are equivalent to their valuations, and then define an ambiguous perturbation by considering ε -contaminations of the initial non-ambiguous beliefs. Thus, in their perturbed ambiguous models it is both common knowledge that types are equivalent to valuations, and that the ambiguous beliefs of each valuation type are induced by an ε -contamination of his initial beliefs, with all higher-order beliefs being trivial.

An ambiguous type space (T^τ, γ^τ) derived from maps τ_i as described above satisfies the definition of an ambiguous perturbation from Sect. 3.3, with $\alpha_i := \tau_i$, and a corresponding precision level given by

$$\varepsilon^\tau := \max_i \sup_{s_i} \bar{d}(\beta_i(s_i), \beta_i(\tau_i(s_i))),$$

where $\beta_i(\tau_i(s_i))$ is the belief hierarchy in $\bar{T}_i$ induced by $\gamma^\tau(\tau_i(s_i))$. Furthermore, the convergence of such perturbations to the initial belief structure can be characterized analogously to the results derived for partitional type spaces:

Proposition 8 *Let (τ_i^n) denote a sequence of maps $\tau_i^n : S_i \rightarrow 2^{\Delta(\Theta \times S_{-i})}$ for each i , such that every τ_i^n is either compact-valued and continuous, or has finite range, and suppose that*

$$\lim_{n \rightarrow \infty} \max_i \sup_{s_i} \sup_{p_i \in \tau_i^n(s_i)} d^P(\eta_i(s_i), p_i) = 0.$$

Then for each $k \geq 1$, $\beta_{ik}(\tau_i^n(s_i))$ converges to $\beta_{ik}(s_i)$ uniformly on S_i in the Hausdorff metric on $\mathcal{H} \Delta X_i^{k-1}$, and hence, the precision levels associated to the ambiguous perturbations defined by the maps τ_i^n converge to zero in n .

Proof For each n , define a set of *virtual types* $S_i^{n*} \subset \Delta(\Theta \times S_{-i})$ by letting

$$S_i^{n*} := \eta_i(S_i) \cup \text{cl} \left\{ \bigcup_{s_i \in S_i} \tau_i^n(s_i) \right\},$$

and extend the belief map η_i to $S_i^* := \bigcup_n S_i^{n*}$ by defining $\eta_i(s_i^*) := s_i^*$ for every $s_i^* \in S_i^{n*}$. Thus, we identify S_i with its image under η_i in $\Delta(\Theta \times S_{-i})$, and view the resulting set as a subset of S_i^* . Next, let T_i^n denote the ambiguous types induced by τ_i^n , and define a map $\eta_i^n : S_i^* \rightarrow \Delta(\Theta \times T_{-i}^n)$ by $\eta_i^n := \mathcal{L}_{(\text{id}, \tau_{-i}^n)} \circ \eta_i$. Then, if $\gamma_i^n := \gamma_i^{\tau^n}$, we have

$$\gamma_i^n(\tau_i^n(s_i)) = \text{cl}\{\eta_i^n(s_i^*) \mid s_i^* \in \tau_i^n(s_i)\},$$

and, as for the case of partitional type spaces, we can define maps $\tilde{\beta}_{ik}^n := \mathcal{L}_{(\text{id}, \beta_{-i}^{k-1})} \circ \eta_i^n$ and $\tilde{\beta}_{ik} := \mathcal{L}_{(\text{id}, \beta_{-i}^{k-1})} \circ \eta_i$, mapping S_i^* to ΔX_i^{k-1} . Substituting τ_i^n for t_i^n and π_i^n , and taking S_i^* as the domain of $\tilde{\beta}_{ik}^n$ and $\tilde{\beta}_{ik}$, the proofs of Lemma 4, Proposition 5 and Corollary 6 apply unaltered, and yield the desired results. $\square$

Note that Proposition 8 requires that the diameter of $\tau_i^n(s_i)$ converges to zero, which is analogous to the requirement that the mesh of the partitions defining partitional type spaces converges to zero, but it requires in addition that $\tau_i^n(s_i)$ converges to $\eta_i(s_i)$ in the respective Hausdorff metric, which is trivially satisfied in the partitional approach. The convergence characterized by Proposition 8 is also equivalent to convergence of the respective type spaces in the metric $\mathbf{d}$, i.e., $\mathbf{d}((S, \eta), (T^{\tau^n}, \gamma^{\tau^n})) \rightarrow 0$. A proof would follow along the same lines as the proof of Proposition 7.

We show next that in many cases, an ambiguous perturbation defined by a type space (T, γ) and surjective maps $\alpha_i : S_i \rightarrow T_i$, can also be derived as part of the class of perturbations considered in this section, through maps $\tau_i : S_i \rightarrow 2^{\Delta(\Theta \times S_{-i})}$. This is clearly the case when the maps α_i define homeomorphisms between S_i and T_i , in which case the α_i 's have continuous inverses, and the following maps are all continuous:

$$S_i \xrightarrow{\alpha_i} T_i \xrightarrow{\gamma_i} \mathcal{H} \Delta(\Theta \times T_{-i}) \xrightarrow{\mathcal{L}_{(\text{id}, \alpha_{-i}^{-1})} \mathcal{H}} \mathcal{H} \Delta(\Theta \times S_{-i}).$$

Hence, the continuous map $\tau_i := \mathcal{L}_{(\text{id}, \alpha_{-i}^{-1})} \mathcal{H} \circ \gamma_i \circ \alpha_i$ associates to each s_i a unique corresponding (compact) subset of $\Delta(\Theta \times S_{-i})$, which can be interpreted as an ambiguous type equivalent to $\alpha_i(s_i)$. This definition of τ_i requires that the α_i 's have continuous inverses, so it does not apply, for example, when the functions α_i are continuous but not invertible. However, if the type sets T_i are finite, so the corresponding perturbations are also finite and the maps α_i have finite range, it is still possible to define the associated perturbation using appropriate maps τ_i . To do that, let (T, γ) and $(\alpha_i)_{i \in N}$ define a finite ambiguous perturbation of a type space (S, η) . Consider for any player

j , the correspondence $\alpha_j^{-1} : T_j \rightrightarrows S_j$,¹⁹ and choose a selection σ_j from α_j^{-1} . Note that the image of T_j under σ_j is a finite subset of S_j that is homeomorphic to the finite set T_j . Since each T_j is finite, the σ_j are continuous, and we can define a map $\tau_i := \mathcal{L}_{(\text{id}, \sigma_{-i})} \circ \gamma_i \circ \alpha_i$ that associates to each s_i a corresponding subset of $\Delta(\Theta \times S_{-i})$, such that $\tau_i(s_i) = \tau_i(s'_i)$ whenever $\alpha_i(s_i) = \alpha_i(s'_i)$. The image of S_i under τ_i is a finite set that is also homeomorphic to T_i . Moreover, under this homeomorphism, the distributions $\gamma_i(\alpha_i(s_i))$ and $\gamma_i^\tau(\tau_i(s_i))$ are equivalent. This construction shows that every finite ambiguous perturbation can be derived as part of the class of perturbations analyzed in this section. The maps τ_i used for this purpose are not necessarily unique, as a consequence of the fact that the type sets S_i can be of higher cardinality than the type sets T_i defining the corresponding finite perturbations.

Proposition 9 *Let (T, γ) be a finite ambiguous ε -perturbation of (S, η) . Then there exist maps $\tau_i : S_i \rightarrow 2^{\Delta(\Theta \times S_{-i})}$ such that $(T, \gamma) = (T^\tau, \gamma^\tau)$.*

The advantage of deriving ambiguous perturbations using maps τ_i is that the perturbations can be constructed directly based on the non-ambiguous type space (S, η) , without having to consider the corresponding Θ -based ambiguous or non-ambiguous universal type space, which also leads to a simple characterization of convergence to the initial unperturbed belief structure. The previous discussion shows that using such an approach is without loss of generality as long as we are only concerned about finite ambiguous perturbations. It also provides an additional perspective on the fact that partitional perturbations only yield a subset of the set of all finite perturbations—if the map τ_i is derived from a partition, then it must be the case that $\eta_i(s_i) \in \tau_i(s_i)$, which is not necessary in the general case. Note also that the set of finite ambiguous ε -perturbations of a non-ambiguous type space $\mathcal{S} = (S, \eta)$ is strictly contained in the set of all finite ambiguous type spaces $\mathcal{T} = (T, \gamma)$ with $\mathbf{d}(\mathcal{S}, \mathcal{T}) \leq \varepsilon$. Referring to the discussion preceding the definition of an ambiguous ε -perturbation (Definition 3), this follows from the fact that an ambiguous ε -perturbation only requires the existence of one ambiguous type that approximates any non-ambiguous type, whereas if $\mathbf{d}(\mathcal{S}, \mathcal{T}) \leq \varepsilon$, there can be many such types.

In addition to being finite, partitional perturbations of type spaces containing a continuum of types will in general be purely ambiguous, i.e., the approximating ambiguous types will have non-singleton belief sets. The construction of perturbations based on maps $\tau_i : S_i \rightarrow 2^{\Delta(\Theta \times S_{-i})}$ can also be used to show that for any arbitrary Θ -based non-ambiguous type space $\mathcal{S} = (S, \eta)$, there exists a corresponding sequence of finite purely ambiguous perturbations $\mathcal{T}^n = (T^n, \gamma^n)$ such that $\mathbf{d}(\mathcal{S}, \mathcal{T}^n) \rightarrow 0$. To state this result, recall that $\mathbf{T}$ was used to denote the set of all ambiguous or non-ambiguous Θ -based type spaces, and let $\mathbf{S}$ denote the set of all non-ambiguous Θ -based type spaces, viewed as a subset of $\mathbf{T}$. Given our assumption that Θ and all type sets in $\mathbf{T}$ are compact, we have the following:

Proposition 10 *For every $\mathcal{S} \in \mathbf{S}$ there exists a sequence $(\mathcal{T}^n)_n$ of finite ambiguous perturbations of $\mathcal{S}$ such that $\mathcal{T}^n \in \mathbf{T} \setminus \mathbf{S}$ for every n , and $\mathbf{d}(\mathcal{S}, \mathcal{T}^n) \rightarrow 0$.*

¹⁹ I.e., we define $\alpha_j^{-1}(t_j)$ to be equal to the inverse image $\alpha_j^{-1}(\{t_j\})$.

Proof Given a Θ -based non-ambiguous type space $\mathcal{S} = (S, \eta)$, we proceed by defining maps $\tau_i^n : S_i \rightarrow 2^{\Delta(\Theta \times S_{-i})}$ that satisfy the assumptions of Proposition 8. Since Θ and type sets in $\mathcal{S}$ are assumed to be compact, the set of measures $\Delta(\Theta \times S_{-i})$ is also compact in the weak topology, and hence, for each natural number n , there exists a finite collection of open $\frac{1}{n}$ -balls in the corresponding Prohorov metric that covers $\Delta(\Theta \times S_{-i})$. Denote this collection by $\mathbb{B}^n$. To define τ_i^n , we can just choose for each $s_i \in S_i$ an element of $\mathbb{B}^n$ that contains $\eta_i(s_i)$, and let $\tau_i^n(s_i)$ be this element. (If there are multiple elements of $\mathbb{B}^n$ that contain $\eta_i(s_i)$, we can choose one of them according to some procedure. For example, we could order the elements of $\mathbb{B}^n$ and choose the one with the smallest index among those containing $\eta_i(s_i)$.) We can then define $\mathcal{T}^n$ to be the ambiguous perturbation induced by the maps τ_i^n , i.e., $\mathcal{T}^n := (T^{\tau^n}, \gamma^{\tau^n})$. Since the diameter of the elements of $\mathbb{B}^n$ converges to zero as $n \rightarrow \infty$, the assumption of Proposition 8 are satisfied, and the claimed convergence result follows. $\square$

6 Discussion

With a view towards providing a framework that can be used to analyze the robustness to ambiguity of various solution concepts, this paper introduces a notion of an ambiguous perturbation of a precise non-ambiguous type space, based on the standard embedding of types into the ambiguous universal type space. This embedding generates belief hierarchies which provide a complete description of a type's beliefs about the parameters describing the game and his opponents' higher-order beliefs. Based on a newly defined metric topology on the set of all ambiguous type spaces, an ambiguous perturbation is defined as an ambiguous type space whose associated belief hierarchies approximate the belief hierarchies of the initial type space, in the sense that these hierarchies are close in the associated product topology derived from the weak topology on probability measures. This product topology yields a natural topology on belief hierarchies, which seems to capture the limited ability of the players or of the researcher to completely specify an infinite hierarchy of beliefs (see also Weinstein and Yildiz 2007, pp. 371–372). Two simple procedures to construct associated perturbations are introduced, and are used to characterize the convergence of the resulting approximations to the initial unperturbed environment without reference to the induced belief hierarchies. One of these procedures is also shown to yield all finite perturbations as a special case.

The introduced class of perturbations can provide a foundation for a robustness analysis for any solution concept for Bayesian games. As a particular application, Stauber (2011) defines a notion of robustness to ambiguous beliefs for Bayesian Nash equilibria, based on the partitional perturbations analyzed in this paper.²⁰ The definition of robustness requires that a robust equilibrium of the initial game with precise beliefs induces a class of ε -equilibria for every partitional ambiguous type space derived from a partition with small enough mesh. The results of this paper show that the robustness notion in Stauber (2011) is consistent with viewing the associated per-

²⁰ The implications regarding the belief hierarchies associated with such partitional perturbations, as discussed in the present paper, are however not considered in Stauber (2011).

turbations as perturbations of belief hierarchies within the ambiguous universal type space. Alternatively, robustness could be defined by considering all ambiguous perturbations associated to the initial type space, or all finite perturbations. The previous results imply that all robust equilibria corresponding to such definitions would also be robust for an analogous robustness definition that is based only on partitional type spaces. A reverse containment may hold in some environments, but this should depend on the particular environment and solution concept used.

Our results also provide additional insight and clarification into the use of ε -contaminations in the construction of ambiguous approximating type spaces in papers such as [Bose et al. \(2006\)](#) and [Bose and Daripa \(2009\)](#). In their auction environment, players' valuations v_i uniquely determine their precise beliefs over the valuations of their opponents, and perturbed ambiguous beliefs are given by ε -contaminations of initial beliefs. Thus, in the resulting perturbation, the same valuations v_i now represent ambiguous types whose beliefs about the (ambiguous) types of their opponents are given by an ε -contamination of their original precise beliefs. Since these perturbations are a special case of the generalized perturbations we consider in Sect. 5, our approach indicates how the analysis in these papers could be extended to more general environments with arbitrary type spaces and ambiguous perturbations.

The literature on robustness to ambiguity in economics has mostly relied on ε -contaminations or metric neighborhoods to define ambiguous perturbations of beliefs. As noted before, these approaches are problematic in a Bayesian game setting. However, many applications in economics, such as the previously mentioned papers by [Bergemann and Schlag \(2011\)](#), [Nishimura and Ozaki \(2004\)](#) and [Lopomo et al. \(2009\)](#), only consider one-sided incomplete information, in which case the problem associated with interacting beliefs does not arise. Alternatively, [Bose et al. \(2006\)](#) and [Bose and Daripa \(2009\)](#) make the standard assumption that valuations uniquely determine a player's precise or ambiguous beliefs about other players' valuations, and that this is common knowledge, so higher-order beliefs are trivial. For a general model described by a type space, our constructions of ambiguous perturbations yield an alternative approach to perturb belief structures that avoids any inconsistency between the treatment of a player's own beliefs and his beliefs about his opponents' beliefs, and does not require strong common knowledge assumptions.

Acknowledgments I thank two anonymous referees for very helpful suggestions. Comments of seminar audiences at the University of Queensland, the 2011 LACEA-LAMES conference, the 2012 SAET conference, and the Fourth Congress of the Game Theory Society, are also gratefully acknowledged.

References

- Ahn DS (2007) Hierarchies of ambiguous beliefs. *J Econ Theory* 136(1):286–301
- Aubin JP, Frankowska H (1990) Set-valued analysis. Birkhäuser
- Bergemann D, Schlag K (2011) Robust monopoly pricing. *J Econ Theory* 146(6):2527–2543
- Berger JO (1985) Statistical decision theory and Bayesian analysis, 2nd edn., Springer series in statistics. Springer, Berlin
- Bewley TF (2002) Knightian decision theory. Part I. *Decis Econ Finance* 25(1):79–110, [Discussion Paper No. 807, Cowles Foundation at Yale University, November 1986].

- Bose S, Daripa A (2009) A dynamic mechanism and surplus extraction under ambiguity. *J Econ Theory* 144(5):2084–2114
- Bose S, Ozdenoren E, Pape A (2006) Optimal auctions with ambiguity. *Theor Econ* 1(4):411–438
- Brandenburger A, Dekel E (1993) Hierarchies of beliefs and common knowledge. *J Econ Theory* 59(1):189–198
- Chassang S, Takahashi S (2011) Robustness to incomplete information in repeated games. *Theor Econ* 6(1):49–93
- Dekel E, Fudenberg D, Morris S (2006) Topologies on types. *Theor Econ* 1(3):275–309
- Dekel E, Fudenberg D, Morris S (2007) Interim correlated rationalizability. *Theor Econ* 2(1):15–40
- Dudley RM (2002) Real analysis and probability. Cambridge studies in advanced mathematics 74. Cambridge University Press, Cambridge
- Friedenberg A, Meier M (2011) On the relationship between hierarchy and type morphisms. *Econ Theory* 46(3):377–399
- Friedenberg A, Meier M (2012) The context of the game. Arizona State University, Working paper
- Gilboa I, Schmeidler D (1989) Maxmin expected utility with non-unique prior. *J Math Econ* 18(2):141–153
- Haimanko O, Kajii A (2012) Approximate robustness of equilibrium to incomplete information. Kyoto University, Working paper
- Harsanyi JC (1967–1968) Games with incomplete information played by Bayesian players, I–III. *Manage Sci* 14:159–182, 320–334, 486–502
- Heifetz A, Samet D (1998) Topology-free typology of beliefs. *J Econ Theory* 82(2):324–341
- Huber PJ (1981) Robust statistics. Wiley, New York
- Kajii A, Morris S (1997) The robustness of equilibria to incomplete information. *Econometrica* 65(6):1283–1309
- Kajii A, Ui T (2005) Incomplete information games with multiple priors. *Jpn Econ Rev* 56(3):332–351
- Kechris AS (1995) Classical descriptive set theory. Graduate texts in mathematics, 156. Springer, Berlin
- Kopylov I (2008) Subjective probability and confidence. Working paper, UC Irvine
- Liu Q (2009) On redundant types and Bayesian formulation of incomplete information. *J Econ Theory* 144(5):2115–2145
- Lopomo G, Rigotti L, Shannon C (2009) Uncertainty in mechanism design. Duke University and UC Berkeley, Working paper
- Mertens JF, Zamir S (1985) Formulation of Bayesian analysis for games with incomplete information. *Int J Game Theory* 14(1):1–29
- Nadler SB (1978) Hyperspaces of sets. Marcel Dekker, New York
- Nishimura KG, Ozaki H (2004) Search and Knightian uncertainty. *J Econ Theory* 119(2):299–333
- Nishimura KG, Ozaki H (2006) An axiomatic approach to ε -contamination. *Econ Theory* 27(2):333–340
- Oyama D, Tercieux O (2010) Robust equilibria under non-common priors. *J Econ Theory* 145(2):752–784
- Penta A (2012) Higher order uncertainty and information: static and dynamic games. *Econometrica* 80(2):631–660
- Stauber R (2011) Knightian games and robustness to ambiguity. *J Econ Theory* 146(1):248–274
- Wasserman LA, Kadane JB (1990) Bayes' theorem for Choquet capacities. *Ann Stat* 18(3):1328–1339
- Weinstein J, Yildiz M (2007) A structure theorem for rationalizability with applications to robust predictions of refinements. *Econometrica* 75(2):365–400